31st October 2016. Vol.92. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A NOVEL TOP-K INFREQUENT MINING TECHNIQUE ON COMPLEX DISTRIBUTED MARKET DATABASES

[1]SUJATHA KAMEPALLI, [2]RAJA SEKHARA RAO KURRA, [3]SUNDARA KRISHNA .Y.K.

[1]Research Scholar,CSE Department,Krishna University,Machilipatnam, Andhra Pradesh,INDIA.

[2]Dean, Sri Prakash College of Engineering,Thuni, Andhra Pradesh,INDIA.

[3] Professor ,CSE Department, Krishna University,Machilipatnam, Andhra Pradesh,INDIA.

E-mail: [1]sujatha101012@gmail.com [2] krr_it@yahoo.co.in [3] yksk2010@gmail.com

ABSTRACT

Infrequent association rule mining is one of the essential tasks in data mining research to find rare items on complex data set. Also, most of the traditional models focus on finding negative association rules based on different association measures. However, finding relational infrequent patterns from the large number of candidate sets is still an open problem in the distributed market analysis. Traditional infrequent mining models are mainly depending on quantitative attributes, limited data size and Boolean datasets. In any distributed environment, as the size and complexity of the market data increases, it is difficult to find the sparsity issue from the positive association rules. In this proposed approach a novel infrequent association mining algorithm was implemented to find the topmost relational infrequent patterns from the complex market dataset. Experimental outcomes prove that the proposed model extracts high quality, infrequent patterns compared to conventional infrequent rule mining techniques.

Keywords: *Complex Data, Infrequent Association Rules, Data Sparsity, Quantitative Association Rules, Rank Correlation.*

1. INTRODUCTION

Association rule mining is an approach to find the hidden patterns and strong relationship patterns in complex datasets. All the generated frequent patterns are not very efficient due to the redundancy in the antecedent or consequent portion of the association rule. In association rule mining, two measures should be used for pattern evaluation i.e support and confidence measures. Practically, infrequent patterns are the complement to the frequent patterns and the infrequent items are too big to be mined in real time applications. In fact, how to discover the infrequent patterns is still a challenging issue in distributed environments. Compared with weighted association rule mining, current research mainly interested to find infrequent patterns by updating incremental rules on complex databases[1].

Given a weighted item set I and a set of transactions as T, we say W has support computation in T, then weighted support of an item set I in a dataset T, denoted as

$$sup_w(I) = count(I)/|T| * W(I);$$

where count(I) is the number of transaction in the set T.

Infrequent Association rule mining was introduced to find associative patterns from market basket data. The market basket data consists of transactions where a transaction is a set of items purchased by a customer. The motivation for applying this data mining approach on market basket data was to learn about buying patterns and use that information in catalog design, and store layout design. Since then, association rule mining has been studied and applied in many other domains (e.g. network intrusion detection, credit card fraud, genetic data analysis). In every domain, there is a need to analyse data to identify patterns associating different attributes. Association rule mining addresses this need.

All the frequent item sets generated from the item set generation step are used to generate maximal frequent item sets. A maximal frequent item set is an item set that is not a subset of any other item set. This stage reduces the number of item sets we are working with significantly. Using the maximal item sets, we generate all the subsets of the maximal item sets and determine if each subset has support counted. As we know from the item set generation stage, some item sets may not be generated because of the item set pruning step and therefore will not have their support counted. Those item sets without support will need to have their support counted prior to the rule generation stage.

Journal of Theoretical and Applied Information Technology
31st October 2016. Vol.92. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

When generating rules from the maximal item sets, a pruned item set may appear on the antecedent or consequent of a rule[2].

In the case of generating association rules, all possible splits of a frequent item set into antecedents and consequents are considered. However, in generating classification based association rules, syntactic constraints are placed such that the rule generated has only items from the classification attribute on the consequent side, while no items from the classification attribute are present on the antecedent side. Further, confidence is calculated for each rule and those rules with confidence greater than or equal to the minimum confidence will form the final set of classification association rules[3].

The choice of minimum support and confidence measures decides , whether association patterns are useful. Most of the traditional association rule mining models only consider a single minimum support for frequent patterns. But many frequent patterns related to items are generated with high support measure. On the other hand, as the size of the support measure increases, we will miss essential patterns related to items with low support measure. To solve this issue, we proposed a new computational measure to discover infrequent patterns.

The main contribution of this paper is to design and develop a novel infrequent mining association model for mining the frequent using the CPTree and infrequent patterns with the modified pruning measure oin the large market data. The rest of the paper is organized as follows. The definition and purpose of frequent association rule mining models are given in section II. Proposed infrequent mining model is described in section III. The experimental results and its analysis are described in the Section IV, and conclusion with future scope is made in Section V.

2.RELATED WORK

The purpose of the association mining is to discover certain patterns between the set of items in the dataset. Infrequent association rules are very essential to many domains, particularly in market and business analysis. Several models have been implemented on infrequent association rule mining, for instance, chi squared models and correlation based models can be used to find the items relationships between the two variables. Association rules decoupled from coherent rules can be reasoned in terms of proportional error reduction [3-6]. All association rules decoupled from coherent rules possess the ability to predict a consequence item set better than the marginal probability. However, not all association rules are decoupled from coherent rules, and so some association rules lack this property. For example, an association rule $X \Rightarrow Y$ with confidence 60% is weaker than guessing that item set Y will be associated by withy item set because it occurs 100% of the time with marginal probability of 1 in a dataset. In contrast, association rules derived from coherent rules will have higher confidence than the marginal probability of item set Y. Otherwise, they are not coherent rules. The concept of coherent rules is independent from any background knowledge such as a specific understanding of the application domain[4]. Suppose T is a set of transaction records. Each record within T will contain a subset of items from the superset I, where I contains all the unique items contained in T. The discovery of association rules from transaction records type of dataset is called Market Basket Analysis. On the other hand, assume D is a set of classification datasets that contains attributes and class attributes. Market Basket Analysis can be used to find the association between the attribute values that are associated with a class. The items in transaction records (T) can be viewed as attributes that have Boolean values. Association rules that are discovered from transaction records (T) are typically called "Boolean Association Rules". On the other hand, attributes in the classification dataset will have richer attribute types [4]. If these attributes have continuous values, they are discretised into several categories and the association rules discovered are called "Quantitative Association Rules".

The pair of rules in equations is a coherent rule that consists of two pseudo-implications of equivalences[5].

(i) $Q1=S(\{i1, \{i7\})$

(ii) $Q2=S(\{i1, \neg\{i7\})$

(iii) $Q3=S(\neg\{i1, \{i7\})$ and

(iv) $Q4=S(\neg\{i1, \neg\{i7\})$

No.	Coherent Rules
1	$\{i1\}\Rightarrow\{i7\}, \neg\{i1\}\Rightarrow\neg\{i7\}$
2	$i1,i3 \Rightarrow\{i7\}, \neg\{i1,i3\}\Rightarrow\neg\{i7\}$
3	$\{i2,i3\}\Rightarrow\{i7\}, \neg\{i2,i3\}\Rightarrow\neg\{i7\}$
4	$\{i4\}\Rightarrow\{i7\}, \neg\{i4\}\Rightarrow\neg\{i7\}$
5	$i3,i4,i5,i6 \Rightarrow\{i7\}, \neg\, i3,i4,i5,i6 \Rightarrow\neg\{i7\}$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Rule generation and rule sorting, same as CP Tree

Associative Classifier Construction [6-8]:

For each rule taken in order from CP Tree

Find the number of remaining examples the rule covers **c**

Find the number of remaining examples the rule correctly classifies **d**

If d is at least 1

Find the percentage **p** of remaining correct classification **c*100/d**

If **p** is at least some threshold

Take the rule in the classifier and remove examples covered by the rule

Else ignore the rule

This classifier ensures that a rule will be included in the final classifier only if it classifies at least 50% of the remaining instances that it covers.

Problems:

Practically this did not work out as it seems theoretically. Accuracy went down. This can be explained from the theoretical background as follows:

· In this process, there is always a chance they over fit the data. A rule that was good in the overall dataset should not be rejected just because some of it's examples have been covered and it has become useless. · If smaller numbers of rules are generated, then selecting rules in this way can leave the dataset uncovered.

In the past decade, a large number of frequent and infrequent mining models have been proposed that include:

[4,9] proposed rule ordering strategies and infrequent mining using weighted measures. These models are applicable to classification based association mining. The main limitation of these models are static weight measure, which are applicable to limited data size.

[7] proposed classification based on predictive rule mining model which combines both the association mining and classification model for pattern analysis. The main problems identified in this model are incorrect class rules and class imbalance .

[6] Proposed a novel inter RBF algorithm which evaluates RBF kernel into Maclaurin series and then discovers frequent and infrequent patterns. Inter RBF is an expention of RBF kernel to learn SVM classifier. The main problem identified in this model is unable to filter large number of infrequent patterns. Also this model generates large number of candidate sets as the size of the attributes increases.

[3] Proposed an efficient high utility sequential pattern miner for frequent and infrequent rules.USpan is the model develop to find frequent and infrequent rules on the sequential databases. This model fails to improve the performance under large datasets with lack of pruning strategies.

[1] Proposed an infrequent wighted itemset mining using FP-growth model. Two novel measures for wighted itemset measure and IWI(infrequent weigthed itemsets) are implemented to discover frequent and infrequent patterns. This model requires domain expert knowledge for patterns analysis and lack of candidate pruning strategies.

[5] Implemented coherent based frequent mining model with support and confidence measures. In this model, a large number of patterns are discovered but misses some intresting infrequent patterns and the pattern's quality.

3. TOP 'K' INFREQUENT MINING ALGORITHM USING RANKING MEASUERES

In this proposed approach, an infrequent mining model was implemented to discover the top frequent rare item sets on the distributed market dataset. The overall workflow is described in the fig 3. For simplicity, the provided data combines and aggregates visited pages from the log files, category and subcategory names, and product related content pages/categories. The aggregate data contain a total of 182 attributes corresponding to pages or categories. This data contains a total of 7296 sessions (each row in the data). For the purpose of market basket analysis, the session data are represented in relational format with unary categorical attributes (a value of "Y" indicates that the corresponding page/category was visited in the session, while a value of "?" indicates that the page/category is missing from the session). In this model, distributed data, minsup are used as input parameters to find the infrequent mining process.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

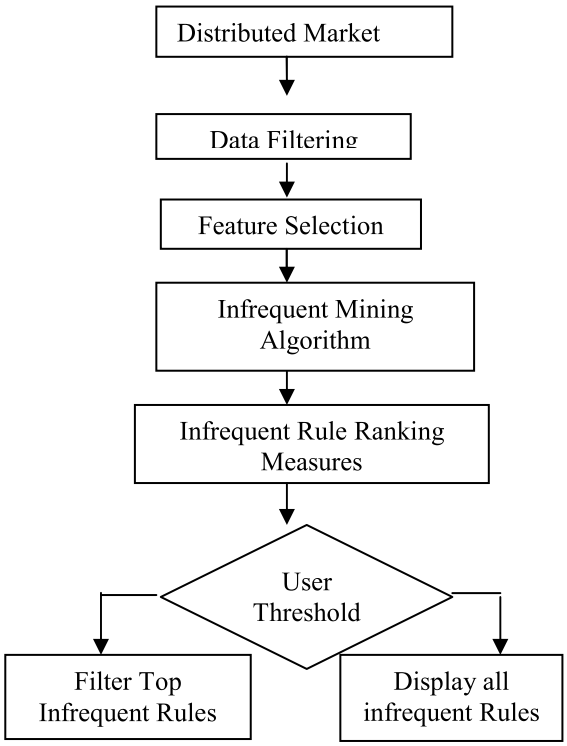

Fig 3. Proposed Flow Chart

Algorithm:
Input: PAS: Positive Association Sets
IAS: Infrequent Association Sets
CS: Candidate Sets

Step 1: Load the distributed market dataset
Step 2: Generate candidate sets using the apriori algorithm.
Step 3: η_{min} minimum weighted threshold

PAS $\leftarrow \varnothing$; IAR $\leftarrow \varnothing$
Find 1-item frequent item set as (I_1)
for (m = 2, I_{m-1} != $\varnothing$, m + +)
do
$CS_m \leftarrow Join\ (I_{m-1}, I_1)$;
done

Step 5: For each item set $t_i \in CS_m$
do
w_i =Find_tfidf(D, t_i)
if $w_i \geq \eta_{min}$ then
$fs_m \leftarrow fs_m \cup \{t_i\}$
done
For each item set itm_set in the{ fs_m }
//m=1…size(fs)
Count cls1=0;
Count cls2=0;
For each item itm in itm_set
Do
If($itm \in class(1)$)
cls1++;
Else
cls2++;
If(cls1>cls2)
then
$fs^1{}_m$ .add(itm_set ,cls1);
else
$fs^2{}_m$ .add(itm_set ,cls2);
done
done
φ_1 =generate_associationrules(CPTree($fs^1{}_m$)
φ_2 = generate_associationrules(CPTree($fs^2{}_m$)
For each rule in φ_1
For each rule in φ_2
do

$$r\sigma_{corr} = RankedCorr(\varphi_1, \varphi_2)$$

if $r\sigma_{corr} \geq min\,thres$
then
if conf(φ_1, φ_2) $\geq$ $conf_{\min}$
then
PAR $\leftarrow$ PAR $\cup$ { φ_1, φ_2 }
else if conf(φ_1, φ_2) $\geq conf_{\min}$ and
sup($\neg\varphi_1, \neg\varphi_2$) $\geq \rho_{\min}$ then
IAR $\leftarrow$ IAR $\cup$ { $\neg\varphi_1, \neg\varphi_2$ }
endif
if $r\sigma_{corr} \leq$ - $min\,thres$ then
if conf($\varphi_1, \neg\varphi_k$) $\geq$ $conf_{\min}$ then
IAR $\leftarrow$ IAR $\cup$ { $\varphi_1, \neg\varphi_k$ }
endif
if conf($\neg\varphi_1, \varphi_k$) $\geq$ $conf_{\min}$ then
IAR $\leftarrow$ IAR $\cup$ { $\neg\varphi_1, \varphi_k$ }
endif
endif
(25) Infrequent_Rules $\leftarrow$ IAR
done
done
done

If D and D' are the frequent item sets and infrequent item sets, then the following condition must hold:

Journal of Theoretical and Applied Information Technology
31st October 2016. Vol.92. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$\{(s(D)+s(D'))/(c(D)+c(D'))\} \geq minsup$$

$$\{(s(D).c(D')+c(D').s(D'))/(c(D)+c(D'))\} \geq minsup$$

$$c(D').s(D') \geq minsup * c(D) + minsup * c(D') - c(D') * s(D')$$

$$c(D')(s(D'-minsup) \geq c(D) * (minsup - s(D))$$

Since $c(D') > 0$ and $s(D) < minsup$ or $minsup - s(D) > 0$, $s(D') - minsup > 0$, weget

$$s(D') > minsup$$

Computational Measures

Lift calculates the ratio between the rules support and confidence of the item set in the rule consequent based on the each selected class.

$$lift = pr(\text{itm}_i / D_i) / pr(\text{itm}_i, D)$$

$pr(\text{itm}_i / D_i)$: Probability of occurrence of an item in samples of ith class.

$pr(\text{itm}_i, D)$: Probability of occurrence of an item in a dataset of ith class.

Correlation(ϕ_1, ϕ_2)=

$$|D_i|\,\text{lift}(i/i \in \phi_1, \phi_1) - |D_i|\,\text{lift}(i \in \phi_2, \phi_2) / |D| \sqrt{\text{lift}(i/i \in \phi_1, \phi_1)^2 - \text{lift}(i/i \in \phi_2, \phi_2)^2}$$

$$|\sum_{t \in \varphi_1} find\_tfid(\varphi_1)| * (sup((A \cup C)_1 U(B \cup D)) - sup(\varphi_1).sup(\varphi_2)$$

$$RankedCorr(\varphi_1(A->B), \varphi_2)) = \frac{\qquad}{\sqrt{sup(A->B).sup(C->D)}}$$

$$find\_tfidf(D, t_i) = (1 + log(f(t, D)).log(N/n(t))$$

Where f(t,D)=probability of the item t in D, N is the total number of items and n(t)=is the number of terms

4.EXPERIMENTAL RESULTS

In this model we have used the distributed market dataset from a real e-commerce site, and use association rule mining to perform market basket analysis on the visitor session data. There are two primary types of products sold through the above site, leg care products, and leg ware products. Each category includes various subcategories and individual products from multiple vendors. There is also a separate categorization of products by specialized "Collections" and by "Assortments."

The data collection mechanism, in addition to capturing click stream page-level data, also captures information on the categories, subcategories, assortments, and collections of products accessed in a given session.

Market data:

@relation 'marketdata'
@attribute articles/dpt_shipping.jhtml {Y}
@attribute main/login.jhtml {Y}
@attribute main/registration.jhtml {Y}
@attribute main/search_results.jhtml {Y}
@attribute main/shopping_cart.jhtml {Y}
@attribute 'Category: Cellulite & Other Treatments' {Y}
@attribute 'Category: Footcare' {Y}
@attribute 'Category: Hair Removal' {Y}
@attribute 'Category: Health Supplements' {Y}
@attribute 'Category: Skincare' {Y}
@attribute /Products/Legwear/AmericanEssentials {Y}
@attribute /Products/Legwear/BellyBasics {Y}
@attribute /Products/Legwear/Berkshire {Y}
@attribute /Products/Legwear/Danskin {Y}
@attribute /Products/Legwear/EllenTracy {Y}
@attribute /Products/Legwear/EvanPicone {Y}
@attribute /Products/Legwear/Givenchy {Y}
@attribute /Products/Legwear/GregNorman {Y}
@attribute /Products/Legwear/Hanes {Y}
@attribute /Products/Legwear/HotSox {Y}
@attribute /Products/Legwear/Oroblu {Y}
@attribute /Products/Legwear/RoundTheClock {Y}
@attribute 'Collection: Better Than Bare - Queen' {Y}
@attribute 'Collection: Beyond Bare Collection' {Y}
@attribute 'Collection: Body Gleamers' {Y}
@attribute 'Collection: Body Gleamers Plus' {Y}
@attribute 'Collection: Body Smoothers' {Y}
@attribute 'Collection: Business Sheers' {Y}
@attribute 'Collection: Celebrate' {Y}
@attribute 'Collection: Childrens Dance' {Y}
@attribute 'Collection: Essential Womens Casual Basics' {Y}
@attribute 'Collection: Essential Womens Dress Basics' {Y}
@attribute 'Collection: Essential Womens Novelty Dress' {Y}
@attribute 'Collection: Essential Womens Sport Basics' {Y}
@attribute 'Collection: Fashion First' {Y}
@attribute 'Collection: Girdle at the Top' {Y}
@attribute 'Collection: Greg Norman Golf Collection' {Y}
@attribute 'Collection: Hanes Plus Collection' {Y}

Journal of Theoretical and Applied Information Technology
31st October 2016. Vol.92. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

@attribute 'Collection: Luxury Collection' {Y}
@attribute 'Collection: Mens City Dress' {Y}
@attribute 'Collection: Mens Country Casual' {Y}
@attribute 'Collection: Mens Essential Sport' {Y}
@attribute 'Collection: Mens Patterns and Textures' {Y}
@attribute 'Collection: Mens Solids' {Y}
@attribute 'Collection: Occasions Collection' {Y}
@attribute 'Collection: Oroblu Fashion Line' {Y}
@attribute 'Collection: Oroblu Italian Hosiery' {Y}
@attribute 'Collection: Passion Privee' {Y}
@attribute 'Collection: Pregnancy Survival Kit' {Y}
@attribute 'Collection: Sheer Confidence' {Y}
@attribute 'Collection: Silky - Queen' {Y}
@attribute 'Collection: Specialty Items' {Y}
@attribute 'Collection: Spring/Summer 2000' {Y}
@attribute 'Collection: Teddy Hose' {Y}
@attribute 'Collection: The New Classics' {Y}
@attribute 'Collection: Womens Dance' {Y}
@attribute 'Collection: Womens Fresh Air Collection' {Y}
@attribute 'Collection: Womens Plus Dance' {Y}
@attribute 'Subcategory: Appliances & Tools' {Y}
@attribute 'Subcategory: Bones & Joints' {Y}
@attribute 'Subcategory: Cellulite' {Y}
@attribute 'Subcategory: Energizers' {Y}
@attribute 'Subcategory: Energizers & Relaxers' {Y}
@attribute 'Subcategory: Exfoliators' {Y}
@attribute 'Subcategory: Gift Sets & Special Items' {Y}
@attribute 'Subcategory: Leg Veins & Circulation' {Y}
@attribute 'Subcategory: Massage/Relaxers' {Y}
@attribute 'Subcategory: Moisturizers' {Y}
@attribute 'Subcategory: Razors & Shaving Treatments' {Y}
@attribute 'Subcategory: Spa Wear & Gear' {Y}
@attribute 'Subcategory: Suncare' {Y}
@attribute 'Subcategory: Tools & Implements' {Y}
@attribute 'Subcategory: Varicose/Spider Veins' {Y}
@attribute /Assortments/Main/LifeStyles {Y}
@attribute /Assortments/Main/UniqueBoutiques {Y}
@attribute /Assortments/Main/Brands/LegCare {Y}
@attribute /Assortments/Main/Brands/Legwear {Y}
@attribute /Assortments/Main/Departments/A1_Hosiery {Y}
@attribute /Assortments/Main/Departments/A2_Socks {Y}
@attribute /Assortments/Main/Departments/A3_Bodywear {Y}
@attribute /Assortments/Main/Departments/A4_Legcare {Y}
@attribute /Assortments/Main/LifeStyles/family {Y}
@attribute /Assortments/Main/LifeStyles/InStyle {Y}
@attribute /Assortments/Main/LifeStyles/LegCare {Y}
@attribute /Assortments/Main/LifeStyles/Sport {Y}
@attribute /Assortments/Main/LifeStyles/Work {Y}
@attribute /Assortments/Main/UniqueBoutiques/01_PlusSizes {Y}
@attribute /Assortments/Main/UniqueBoutiques/02_men {Y}
@attribute /Assortments/Main/UniqueBoutiques/03_kids {Y}
@attribute /Assortments/Main/UniqueBoutiques/04_evenings {Y}
@attribute /Assortments/Main/UniqueBoutiques/05_maternity {Y}
@attribute /Assortments/Main/UniqueBoutiques/06_dance {Y}
@attribute /Assortments/Main/UniqueBoutiques/07_gifts {Y}
@attribute /Assortments/Main/UniqueBoutiques/08_Seasonal {Y}

5. EXPERIMENTAL RESULTS EVALUATION:

In this section we presented our experimental results on Market datasets. Following are the infrequent patterns identified using the proposed infrequent pattern mining model. These patterns are filtering using the Lift and correlation measures . In our model, as the size of the combined measure Lift and Correlation increases, topmost infrequent patterns are filtered on the large number of candidate infrequent sets.

Filtered Infrequent Patterns:
[/Products/Legwear=Y, Collection: Beyond Bare Collection=Y]: 704 ==>
[/Assortments/Main/UniqueBoutiques=Y]: 282
[Category: Footcare=Y, Subcategory: Gift Sets & Special Items=Y]: 599 ==>
[/Assortments/Main/Brands/Legwear=Y]: 240
[/Assortments/Main/Brands/Legwear=Y]: 2582 ==>
[/Assortments/Main/UniqueBoutiques=Y]: 1035

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[Category: Hair Removal=Y, /Products/Legwear=Y]: 1177 ==> [/Assortments/Main/UniqueBoutiques=Y]: 472
[/Assortments/Main/Brands/Legwear/Givenchy=Y]: 692 ==> [/Products/Legwear=Y, /Assortments/Main/UniqueBoutiques=Y]: 278
[/Products/Legwear/EvanPicone=Y]: 567 ==> [Category: Footcare=Y, /Products/Legwear=Y]: 228
[/Products/Legwear=Y, /Products/Legwear/EvanPicone=Y]: 567 ==> [Category: Footcare=Y]: 228
[/Products/Legwear/EvanPicone=Y]: 567 ==> [Category: Footcare=Y]: 228
[/Products/Legwear=Y, Subcategory: Cellulite=Y]: 833 ==> [/Assortments/Main/UniqueBoutiques=Y]: 335
[Category: Cellulite & Other Treatments=Y, /Products/Legwear=Y]: 1061 ==> [/Assortments/Main/UniqueBoutiques=Y]: 427
[Category: Skincare=Y]: 1831 ==> [/Assortments/Main/UniqueBoutiques=Y]: 737
[Category: Footcare=Y, /Assortments/Main/Brands/Legwear=Y]: 861 ==> [/Products/Legwear=Y, /Assortments/Main/UniqueBoutiques=Y]: 347
[/Assortments/Main/Brands/Legwear/Oroblu=Y]: 553 ==> [/Products/Legwear=Y, /Products/Legwear/Oroblu=Y, Collection: Oroblu Italian Hosiery=Y]: 223
[/Assortments/Main/Brands/Legwear/Oroblu=Y]: 553 ==> [/Products/Legwear/Oroblu=Y, Collection: Oroblu Italian Hosiery=Y]: 223
[/Assortments/Main/Brands/Legwear/Hanes/AssortmentList/1_silkreflections=Y]: 654 ==> [Category: Footcare=Y]: 264
[Category: Skincare=Y, /Products/Legwear=Y, /Assortments/Main/UniqueBoutiques=Y]: 623 ==> [/Assortments/Main/Brands/Legwear=Y]: 252
[Category: Footcare=Y]: 2124 ==> [/Assortments/Main/Brands/Legwear=Y]: 861
[/Products/Legwear/Danskin=Y, Collection: Womens Dance=Y]: 618 ==> [/Products/Legwear=Y, /Assortments/Main/UniqueBoutiques=Y]: 251
[/Products/Legwear=Y, /Products/Legwear/Danskin=Y, Collection: Womens Dance=Y]: 618 ==> [/Assortments/Main/UniqueBoutiques=Y]: 251
[/Products/Legwear/Danskin=Y, Collection: Womens Dance=Y]: 618 ==> [/Assortments/Main/UniqueBoutiques=Y]: 251
[/Products/Legwear=Y, Collection: Womens Dance=Y]: 646 ==> [/Assortments/Main/UniqueBoutiques=Y]: 263
[Category: Skincare=Y, /Products/Legwear=Y]: 1528 ==> [/Assortments/Main/UniqueBoutiques=Y]: 623
[/Products/Legwear=Y, /Assortments/Main/Brands/Legwear=Y]: 2268 ==> [/Assortments/Main/UniqueBoutiques=Y]: 925
[/Products/Legwear/Danskin=Y]: 1098 ==> [/Products/Legwear=Y, /Assortments/Main/Brands/Legwear=Y]: 448
[/Products/Legwear=Y, /Products/Legwear/Danskin=Y]: 1098 ==> [/Assortments/Main/Brands/Legwear=Y]: 448
[/Products/Legwear/Danskin=Y]: 1098 ==> [/Assortments/Main/Brands/Legwear=Y]: 448
[/Products/Legwear/AmericanEssentials=Y]: 825 ==> [Category: Footcare=Y, /Products/Legwear=Y]: 337
[/Products/Legwear=Y, /Products/Legwear/AmericanEssentials=Y]: 825 ==> [Category: Footcare=Y]: 337
[/Products/Legwear/AmericanEssentials=Y]: 825 ==> [Category: Footcare=Y]: 337
[/Products/Legwear=Y, Subcategory: Bones & Joints=Y]: 580 ==> [Category: Health Supplements=Y, /Assortments/Main/Brands/Legwear=Y]: 237
[Category: Health Supplements=Y, /Products/Legwear=Y, Subcategory: Bones & Joints=Y]: 580 ==> [/Assortments/Main/Brands/Legwear=Y]: 237
[/Products/Legwear=Y, Subcategory: Bones & Joints=Y]: 580 ==> [/Assortments/Main/Brands/Legwear=Y]: 237
[Subcategory: Energizers & Relaxers=Y]: 695 ==> [Category: Footcare=Y, /Assortments/Main/Brands/Legwear=Y]: 284
[Category: Footcare=Y, Subcategory: Energizers & Relaxers=Y]: 695 ==> [/Assortments/Main/Brands/Legwear=Y]: 284
[Subcategory: Energizers & Relaxers=Y]: 695 ==> [/Assortments/Main/Brands/Legwear=Y]: 284
[/Assortments/Main/Brands/Legwear/AmericanEssentials=Y]: 567 ==> [/Products/Legwear=Y, /Products/Legwear/AmericanEssentials=Y]: 232
[/Assortments/Main/Brands/Legwear/AmericanEssentials=Y]: 567 ==> [/Products/Legwear/AmericanEssentials=Y]: 232
[/Products/Legwear=Y, /Assortments/Main/UniqueBoutiques=Y, /Assortments/Main/Departments/A1_Hosiery=Y]: 569 ==> [main/search_results.jhtml=Y]: 233
[Category: Skincare=Y, /Assortments/Main/Brands/Legwear=Y]: 686 ==> [/Assortments/Main/UniqueBoutiques=Y]: 281

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[/Products/Legwear=Y, Collection: Oroblu Italian Hosiery=Y]: 744 ==> [/Products/Legwear/Oroblu=Y, /Assortments/Main/UniqueBoutiques=Y]: 305
[/Products/Legwear=Y, /Assortments/Main/Brands/Legwear/DKNY=Y]: 859 ==> [/Assortments/Main/UniqueBoutiques=Y]: 353
[Category: Cellulite & Other Treatments=Y, /Products/Legwear=Y, Subcategory: Cellulite=Y]: 636 ==> [/Products/Legwear/Hanes=Y]: 262
[Subcategory: Gift Sets & Special Items=Y]: 1046 ==> [/Assortments/Main/UniqueBoutiques=Y]: 432
[/Products/Legwear=Y, /Assortments/Main/Brands/Legwear/Hanes/AssortmentList/1_silkreflections=Y]: 559 ==> [Category: Footcare=Y]: 231
[Category: Footcare=Y, /Products/Legwear=Y]: 1787 ==> [/Assortments/Main/Brands/Legwear=Y]: 740
[/Assortments/Main/Brands/Legwear/Oroblu=Y]: 553 ==> [/Products/Legwear=Y, Collection: Oroblu Italian Hosiery=Y]: 229
[Category: Skincare=Y, /Products/Legwear=Y, /Assortments/Main/Brands/Legwear=Y]: 608 ==> [/Assortments/Main/UniqueBoutiques=Y]: 252
[Category: Cellulite & Other Treatments=Y, Subcategory: Cellulite=Y]: 774 ==> [/Assortments/Main/UniqueBoutiques=Y]: 321
[/Assortments/Main/Departments/A1_Hosiery=Y]: 1371 ==> [/Products/Legwear=Y, /Assortments/Main/UniqueBoutiques=Y]: 569
[/Products/Legwear=Y, Collection: Oroblu Italian Hosiery=Y]: 744 ==> [/Assortments/Main/Brands/Legwear=Y]: 309
[Subcategory: Spa Wear & Gear=Y]: 737 ==> [Category: Footcare=Y, /Assortments/Main/UniqueBoutiques=Y]: 307
[Category: Footcare=Y, Subcategory: Spa Wear & Gear=Y]: 737 ==> [/Assortments/Main/UniqueBoutiques=Y]: 307
[Subcategory: Spa Wear & Gear=Y]: 737 ==> [/Assortments/Main/UniqueBoutiques=Y]: 307
[/Products/Legwear/Oroblu=Y, Collection: Oroblu Italian Hosiery=Y]: 680 ==> [/Products/Legwear=Y, /Assortments/Main/Brands/Legwear=Y]: 284
[/Products/Legwear=Y, /Products/Legwear/Oroblu=Y, Collection: Oroblu Italian Hosiery=Y]: 680 ==> [/Assortments/Main/Brands/Legwear=Y]: 284
[/Products/Legwear/Oroblu=Y, Collection: Oroblu Italian Hosiery=Y]: 680 ==> [/Assortments/Main/Brands/Legwear=Y]: 284
[/Assortments/Main/Brands/Legwear/Oroblu=Y]: 553 ==> [Collection: Oroblu Italian Hosiery=Y]: 231
[Category: Health Supplements=Y, /Products/Legwear=Y]: 840 ==> [/Assortments/Main/UniqueBoutiques=Y]: 351
[/Products/Legwear/Oroblu=Y]: 935 ==> [/Products/Legwear=Y, /Assortments/Main/Brands/Legwear=Y]: 391
[/Products/Legwear=Y, /Products/Legwear/Oroblu=Y]: 935 ==> [/Assortments/Main/Brands/Legwear=Y]: 391
[/Products/Legwear/Oroblu=Y]: 935 ==> [/Assortments/Main/Brands/Legwear=Y]: 391
[/Assortments/Main/Brands/Legwear/Givenchy=Y]: 692 ==> [/Assortments/Main/UniqueBoutiques=Y]: 290
[/Assortments/Main/Brands/Legwear/SilkReflections=Y]: 539 ==> [/Assortments/Main/Brands/Legwear=Y]: 226
[/Products/Legwear=Y, Subcategory: Spa Wear & Gear=Y]: 627 ==> [Category: Footcare=Y, /Assortments/Main/UniqueBoutiques=Y]: 263
[Category: Footcare=Y, /Products/Legwear=Y, Subcategory: Spa Wear & Gear=Y]: 627 ==> [/Assortments/Main/UniqueBoutiques=Y]: 263
[/Products/Legwear=Y, Subcategory: Spa Wear & Gear=Y]: 627 ==> [/Assortments/Main/UniqueBoutiques=Y]: 263
[/Assortments/Main/Brands/Legwear/Hanes=Y]: 922 ==> [/Products/Legwear=Y, /Products/Legwear/Hanes=Y]: 387
[/Assortments/Main/Brands/Legwear/Hanes=Y]: 922 ==> [/Products/Legwear/Hanes=Y]: 387
[/Assortments/Main/Brands/Legwear/Hanes=Y]: 922 ==> [/Products/Legwear=Y, /Assortments/Main/Brands/Legwear=Y]: 388
[/Products/LegCare=Y, /Assortments/Main/UniqueBoutiques=Y]: 726 ==> [/Products/Legwear=Y]: 306
[/Assortments/Main/UniqueBoutiques=Y, /Assortments/Main/Departments/A1_Hosiery=Y]: 633 ==> [/Assortments/Main/Brands/Legwear=Y]: 267
[/Assortments/Main/Brands/Legwear/DonnaKaran=Y]: 692 ==> [/Products/Legwear=Y, /Assortments/Main/Brands/Legwear=Y]: 292
[/Products/Legwear=Y, Subcategory: Gift Sets & Special Items=Y]: 812 ==> [/Assortments/Main/UniqueBoutiques=Y]: 343
[/Assortments/Main/Brands/Legwear/DonnaKaran=Y]: 692 ==> [/Assortments/Main/UniqueBoutiques=Y]: 293
[/Products/Legwear=Y, /Assortments/Main/Brands/Legwear/DonnaKaran=

Journal of Theoretical and Applied Information Technology
31st October 2016. Vol.92. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Y]: 592 ==>
[/Assortments/Main/UniqueBoutiques=Y]: 251
[Category: Cellulite & Other Treatments=Y, /Products/Legwear=Y, Subcategory: Cellulite=Y]: 636 ==> [/Assortments/Main/UniqueBoutiques=Y]: 270
[Collection: Womens Dance=Y]: 716 ==> [/Products/Legwear=Y, /Products/Legwear/Danskin=Y, /Assortments/Main/Brands/Legwear=Y]: 304
[Collection: Womens Dance=Y]: 716 ==> [/Products/Legwear/Danskin=Y, /Assortments/Main/Brands/Legwear=Y]: 304
[/Products/Legwear/Givenchy=Y]: 696 ==> [/Products/Legwear=Y, /Assortments/Main/UniqueBoutiques=Y]: 297
[/Products/Legwear=Y, /Products/Legwear/Givenchy=Y]: 696 ==> [/Assortments/Main/UniqueBoutiques=Y]: 297
[/Products/Legwear/Givenchy=Y]: 696 ==> [/Assortments/Main/UniqueBoutiques=Y]: 297
[/Products/Legwear/EvanPicone=Y]: 567 ==> [/Products/Legwear=Y, /Assortments/Main/UniqueBoutiques=Y]: 242
[/Products/Legwear=Y, /Products/Legwear/EvanPicone=Y]: 567 ==> [/Assortments/Main/UniqueBoutiques=Y]: 242
[/Products/Legwear/EvanPicone=Y]: 567 ==> [/Assortments/Main/UniqueBoutiques=Y]: 242
[/Assortments/Main/Brands/Legwear/AmericanEssentials/AssortmentList/2_menscasualsport=Y]: 665 ==> [Category: Skincare=Y, /Products/Legwear=Y]: 284
[/Products/Legwear=Y, Subcategory: Energizers & Relaxers=Y]: 589 ==> [Category: Footcare=Y, /Assortments/Main/Brands/Legwear=Y]: 252
[Category: Footcare=Y, /Products/Legwear=Y, Subcategory: Energizers & Relaxers=Y]: 589 ==> [/Assortments/Main/Brands/Legwear=Y]: 252
[/Products/Legwear=Y, Subcategory: Energizers & Relaxers=Y]: 589 ==> [/Assortments/Main/Brands/Legwear=Y]: 252
[/Products/Legwear/Danskin=Y]: 1098 ==> [/Products/Legwear=Y, /Assortments/Main/UniqueBoutiques=Y]: 470
[/Products/Legwear=Y, /Products/Legwear/Danskin=Y]: 1098 ==> [/Assortments/Main/UniqueBoutiques=Y]: 470

6. PERFORMANCE ANALYSIS

Our experiment is made on a PC with 3.0GHZ CPU and 8GB memory. Proposed model was implemented in Java programming environment with third party java libraries such as Jama, Apache math , Apache Commons. We conducted experiments to evaluate the performance of proposed model with the traditional models such as CPTree, UPSpan in terms of memory , computation cost and number of infrequent patterns.

Table 1: Computational Time (Ms)

Number of Paterns	CPTree(ms)	UPSpan(ms)	Proposed Model(ms)
100	2524	1435	974
150	4633	2755	1748
200	6278	5297	2548
250	8734	7955	4288
350	9123	8935	6297
500	14644	10465	9143

Table 1, describes the computation time of the proposed infrequent mining model compared to traditional models such as CPTree , Multiclass association model UPSpan . From the table, it is observed that proposed model has less computation time as the number of patterns increases.

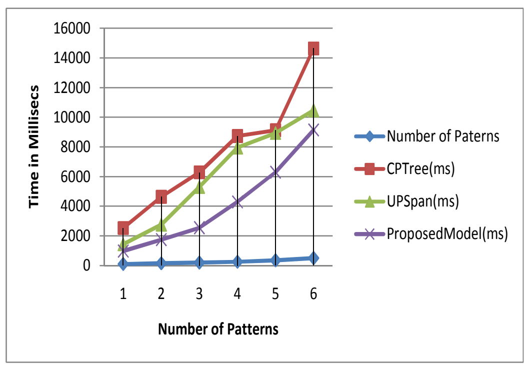

Fig 4: Computational Time In Various Models

Figure 4, describes the computation time of the proposed infrequent mining model compared to traditional models such as CPTree , Multiclass association model UPSpan . From the figure, it was observed that proposed model has less computation time as the number of patterns increases.

Journal of Theoretical and Applied Information Technology
31st October 2016. Vol.92. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

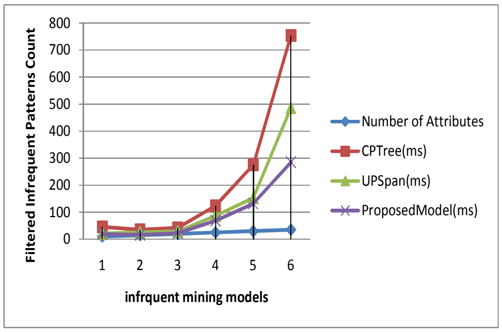

Fig 5: Filtered Infrequent Rules Using Attributes Size

Figure 5 , describes the number of filtered infrequent patterns on the total number of attributes. From the Figure, it was observed that proposed infrequent mining model has topmost filtered patterns compared to traditional models such as CPTree , Multiclass association model UPSpan .

Table 2: Filtered Infrequent Rules In Traditional And Proposed Model

Number of Attributes	CPTree(ms)	UPSpan(ms)	Proposed Model(ms)
10	46	21	19
15	35	26	17
20	42	29	23
25	124	87	69
30	275	153	132
35	754	486	285

Table 1, describes the number of filtered infrequent patterns on the total number of attributes. From the table, it was observed that proposed infrequent mining model has topmost filtered patterns compared to traditional models such as CPTree , Multiclass association model UPSpan .

7.CONCLUSION AND FUTURE SCOPE

In this proposed work, an optimized infrequent mining algorithm was implemented on the real time distributed market dataset. Due to the sparsity and inefficient pruning problems, traditional models such as a coherent mining model, CP-tree model and FP-tree model are not suitable to find the infrequent patterns from the large collection of positive frequent rules. In any distributed environment, as the size and complexity of the market data increases, it is difficult to find the sparsity issue from the positive association rules. In our proposed approach a novel infrequent association mining algorithm was implemented to find the topmost relational infrequent patterns from the complex market dataset. Our model has three main achievement : 1. Applicable to mixed attributes with large number of dimensions. 2. When large number of itemsets are produced , our proposed measures are taken into account as filter criterion. 3. This model doesn't require any domain knowledge and will not miss any item with low threshold. Experimental outcomes prove that the proposed model extracts high quality, infrequent patterns compared to conventional infrequent rule mining techniques. The main limitations of our research model are:

1) As the size of the database increases, requires high computation memory to filter infrequent patterns.
2) This model is not applicable to streaming data.

In future , this work can be extended to dynamic streaming data and hadoop framework.

REFERENCES

[1]“Cagliero,luca, Garza, paolo,”Infrequent weighted itemset mining using frequent pattern growth”, IEEE transactions on knowledge and data engineering, vol 26, 2014.

[2]Wenjuan dong, he jiang, guoling liu,”Incremental updating algoritm for infrequent itemsets on weighted condition”, IEEE ,2010.

[3]Junfu Yin, Zhigang Zheng, Longbing Cao,"USpan: an efficient algorithm for mining high utility sequential patterns",KDD '12: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining,2012.

[4]Ling Zhou, Stephen Yau,"Association rule and quantitative association rule mining among infrequent items"MDM '07: Proceedings of the 8th international workshop on Multimedia data mining: (associated with the ACM SIGKDD 2007).

[5]Aditya A. Davale; Shailendra W. Shende,"Implementation of coherent rule mining algorithm for association rule mining",Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), 2015 International Conference on 2015.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[6]Quanzhong Liu; Yang Zhang; Zhengguo Hu,"Extracting Positive and Negative Association Classification Rules from RBF Kernel",IEEE,2010"

[7]Zhixin Hao; Xuan Wang; Lin Yao; Yaoyun Zhang,"Improved classification based on predictive association rules",IEEE,2009.

[8]Nitendra Kumar Vishwakarma; Jatin Agarwal; Sankalp Agarwal; Shantanu Sharma," Comparative analysis of different techniques in classification based on association rules",IEEE,2013

[9]Yanbo J. Wang; Qin Xin; Frans Coenen ,"A Novel Rule Weighting Approach in Classification Association Rule Mining",IEEE,2007.